DIALOG AUGMENTATION & INTENT CLASSIFICATION

Aseem Srivastava
IIIT Delhi / Lab LCS2
aseems@iiitd.ac.in

Abstract

Intent Detection in Dialogue System is a decisive step to drive the conversation in dialogue system in the right direction. Often, Language Understanding part is paid less attention while building a dialogue system and allows ambiguous response generation. We will try to reconstruct the well-known problem statement of Intent Detection for minority intents and augment dialogues with similar intent in order to create better understanding of minority intents. Proposed architecture and mechanism shows good results on dialogues and can the particular augmentation mechanism could also be extended to other dialogue related classification problems. In this research it work will generate better understanding with even lesser number of intent classifiers using cutting edge machine learning models and state of the art deep learning techniques on goal oriented dialogue systems.

1 Introduction

Intent Detection(ID) in Conversational Dialogues is a very crucial phase while revolutionizing dialogue systems with Machine Learning. Leveraging the Natural Language Understanding(LU), Intent Detection helps in improvising the language generation and advancement in Dialogue Management. Previously, the task of intent detection has been done many times with different datasets. Here we will try to detect the indents precisely where the number of training examples is very less in number and encounter the phase where chatbots are unable to figure out the intent of the user.

For dialogue systems, especially for goal-oriented dialogue systems, a chatbot should be robust enough to detect the intent of the user without taking any particular accent into assumption. And that is where we are dealing even with a lesser number of intents classified in the dataset and bring them to correct identification of indents while in the test phase. For the purpose, we will be using the dataset released by Sonos in June 2017. We will see more about the data in the Dataset section.

We will look into paraphrasing techniques to leverage the minority intent class and augment similar dialogues in order to have better understanding of the dialogues in a dialog system.

2 Importance

Dialogue Systems are build with the focus to complete the task or domain on which they are trained on. Generally, researchers pay less attention to the LU with Intents part to diagonose the intents with very less occurrence in the conversation. The ID with the help of various machine learning techniques would help us to create a better understanding of human intents and flow of conversation.

At the same time our output would also create a positive impact on the dialog management system to reproduce better language generation models.

According to the information extracted, the system can then decide on the appropriate actions to be taken, to help the users achieve their demands. LU applications are becoming increasingly significant in our everyday lives. Numerous devices, such as smartphones, have personal assistants that are built with LU technologies. Multiple product support systems like help centers use ID to reduce the need for a large number of employees that copy-and-paste boring responses to frequently asked questions. Chat bots, automated email responders, answer recommenders from a knowledge base with questions and answers.

Intent	Dialogue
atis_flight	i want to fly from boston at 838 am and arrive in denver at 1110 in the morning
atis_flight_time	what is the arrival time in san francisco for the 755 am flight leaving washington

Table 1: Above table shows example dialogues from user to the bot and corresponding annotated intents in the ATIS dataset.

3 Dataset

The dataset used for ID task is ATIS - Airline Travel Information System which is a benchmark dataset originally released by Microsoft. In the dataset we are dealing with 8 different intents related to airline travel and flight scheduling protocols. Table 2 shows the tabular view of intents in the dataset along with their frequency.

Intent	Frequency
atis_abbreviation	147
atis_aircraft	81
atis_airfare	423
atis_airline	157
atis_flight	3666
atis_flight_time	54
atis_ground_service	255
atis_quantity	51

Table 2: Intents present in the ATIS dataset and their corresponding frequency.

Also, ATIS dataset contains 4,978 training samples along with a vocabulary size of 943. The benchmark dataset is commonly used for slot filing problem statement too providing 129 slot counts. The part in which we are interested in is intent count which are 26 in total. For our experiment we are dealing with 8 different intents related to airline travel and flight scheduling protocols. Table 2 shows the tabular view of intents and their counts.

4 Baseline Model Selection

We first briefly describe the BERT model (Devlinet al., 2018) and then introduce the proposed joint model based on BERT. Figure 2 illustrates a high-level view of the proposed model.

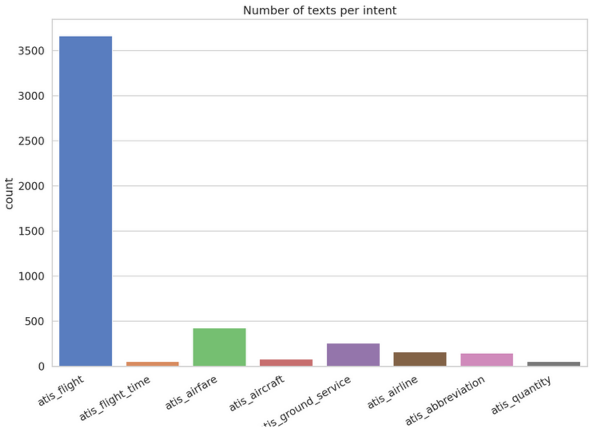

Figure 1: Intent frequency count chart

Since our target is to encounter the minority class we will dive deep into the Siamese network.

4.1 BERT

The model architecture of BERT is a multi-layer bidirectional Transformer encoder based on the original Transformer model (Vaswani et al., 2017). The input representation is a concatenation of Word Piece embeddings (Wu et al., 2016), positional embeddings, and the segment embedding. Specially, for single sentence classification and tagging tasks, the segment embedding has no discrimination.

A special classification embedding([CLS]) is inserted as the first token and a special token ([SEP]) is added as the final token. Given an input token sequence x= (x1, . . . , xT), the output of BERT is H= (h1, . . . ,hT). The BERT model is pre-trained with two strategies on large-scale unlabeled text, i.e., masked language model and next sentence prediction.The pre-trained BERT model provides a powerful context-dependent sentence representation and can be used for various target tasks, i.e., intent classification, through the fine-tuning procedure, similar to how it is used for other NLP tasks

4.1.1 Training Details

We use English uncased BERT-Base model1,which has 12 layers, 768 hidden states, and 12heads.BERT is pretrained on Books Corpus(800M words) (Zhu et al., 2015) and English Wikipedia (2,500M words). For fine-tuning, all hyper-parameters are tuned on the development set. The maximum length is 50. The batch size is128. Adam (Kingma and Ba, 2014) is used for

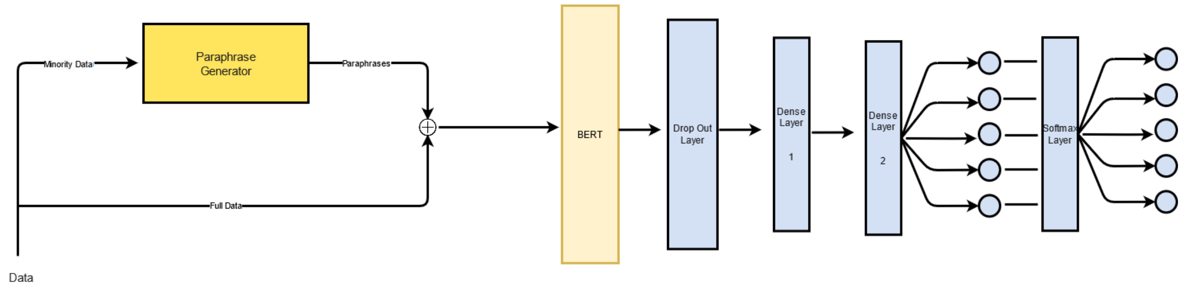

Figure 2: Architecture of the proposed model which is further divided into two components: (i) The paraphrase generation component. (ii) The classifier feed forward neural network. Output received is the softmaxed probability distribution of intent classes

optimization with an initial learning rate of 5e-5. The dropout probability is 0.1. The maximum number of epochs is selected from [1, 5, 10, 20, 30, 40].

5 Our Approach

Our Approach is a mix of two different components where at first dialogues with minority intent are leveraged by augmenting paraphrases of the dialogue utterances. Since these utterances belong to the same intent as that of the parent dialogue, ATIS dataset can be augmented to three times. In our work we augmented data for dialogues from 4 minor intent-classes. For paraphrase generation we used a pretrained T5 model. Secondly along with these dialogues, a neural network is trained from scratch on top of the BERT embeddings.

In our approach we used the BERT based encoder part of transformer architecture combined with our implementation of feed forward neural network. The whole architecture is shown in Figure2.

As an input we used BERT embeddings pretrained on unsupervised corpus. Those inputs were fed to a feed forward neural network (Figure2). This module consisted of a single drop layer followed by two feed-forward layer with hidden dimension of 512.

5.1 Paraphrase Generation

To deal with limited data problem, we introduced novel data augmentation integration with intent detection problem for the first time. Our paraphrase generation model uses a T5 model which is robust to many NLP tasks. T5 is a new transformer model from Google that is trained in an end-to-end manner with text as input and modified text as output. T5 model achieves state-of-the-art results on many NLP problem statements like summarization, sentence extraction, machine translation, etc using the same text-to-text transformer trained on a huge text corpus.

In this project we trained T5 with the original sentence as input and paraphrased (duplicate sentence for ATIS dataset) sentence as output.

5.2 Intent Classifier

We used BERT embeddings to pass from a 3 layer feed forward neural network giving 5 outputs which are at last passed through a softmax layer.

Our feed forward neural network contains a drop layer at the start with dropout value of 0.5 on top of which ReLU activation function is applied. The output from this activation function is passed through first linear layer taking 768 dimensions as input and producing 512 dimensional output to the next hidden layer. Now the second linear layer produces 5 outputs for classification of intents which is further passed through a softmax function in order to get probability distribution. The training of model on regular 8GB RAM of i5 processor for 30 epochs and 32 batch size took 2 hours for data augmentation and 12 hours to train the model on the classification architecture.

6 Results

We used a regular accuracy metric to compare results with baseline and other models. We also evaluated F1 Score for our model but since other models have not used so it become irrelevant to compare on. However precision, recall, f1 score are all calculated and presented.

Dataset	ATIS Train	ATIS Test
BERT-Softmax (w/ tuning)	75.83%	79.00%
BERT-Softmax (w/o paraphrase w/tuning)	80.90%	81.10%
Dialog Augmented BERT	90.0%	87.13%

Table 3: Performance of the proposed architecture as well as baseline models with metric - accuracy. Moreover these architectures are trained on standard split.

7 Code Repository

All the codes and dataset can be found here at click here

References

[1] Chen, Qian, Zhu Zhuo, and Wen Wang. ”Bert for joint intent classification and slot filling.” arXiv preprint arXiv:1902.10909 (2019).

[2] Dataset: https://github.com/sonos/nlu-benchmark

[3] Firdaus, Mauajama, et al. ”Intent detection for spoken language understanding using a deep ensemble model.” Pacific Rim international conference on artificial intelligence. Springer, Cham, 2018

[4] Coucke, Alice, et al. ”Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces.” arXiv preprint arXiv:1805.10190 (2018).

[5] Niu, Peiqing, Zhongfu Chen, and Meina Song. ”A novel bi-directional interrelated model for joint intent detection and slot filling.” arXiv preprint arXiv:1907.00390 (2019).